

Homework 5

In this homework, you will be writing a short program in Python to find duplicate words in a file. This assignment focuses on Chapter 5, specifically reading data from a file.

Program Features

A common mistake when writing is duplicating a word ("the the"). This is a big enough problem that Microsoft Word and other text editing software will mark these mistakes as errors. In this assignment, we will develop a short Python program that will find these duplicate words and alert the user by printing out a simple message. I suggest using a for loop to read the file data line by line. Once you have a line you will need to perform further string manipulations to determine if there's a duplicate word.

Given the following (slightly mangled) text stored inside of `quote.txt`:

```
He that would make his own liberty liberty secure,
must guard even his enemy from oppression;
for for if he violates this duty, he
he establishes a precedent that will reach to himself.
-- Thomas Paine
```

Running your program would produce the following output (user input is green):

```
Enter file name: quote.txt
Found word: "liberty" on line 1.
Found word: "for" on line 3.
Found word: "he" on line 4.
```

As shown above, your program should also be able to find duplicate words **across lines** in the input file.

Helpful Hints

We haven't gotten to lists yet, but we can make use of one for this assignment without knowing all the details. Here's an example of how to get all of the words in a line and turn them into separate strings:

```
line = "This is a line from a file"
words = line.split() # We saw split() in Chapter 4. It will break up
                     # our string into just the words (no whitespace)
                     # and maintain the order. So the variable words
                     # contains "This", "is", "a" "line", "from",
                     # "a", and "file".

# To access those individual words (separate strings) use a for loop:
for word in words:
    print(word)
```

Each iteration of the for loop will print out a word. The variable `word` becomes "This", then "is" and so on, until all of the words have been printed. The result of the loop is shown below:

```
This
is
a
line
from
a
file
```

What to Submit

For this assignment you should submit your **hw5.py** file.

This assignment will be graded automatically. Test your programs thoroughly before submitting them. Make sure that your programs produce correct results for every logically valid test case you can think of. Do not waste submissions on untested code, or on code that does not run with the supplied code from the course website.

Web-CAT will assign a score based on runtime testing of your submission; your best score will be counted; the TAs will later verify that your best submission meets the stated restrictions, and assess penalties if not.

To submit this assignment:

1. Visit <http://web-cat.cs.vt.edu> in your web browser.
2. Enter your Virginia Tech PID and password in the appropriate fields on the log-in screen, and make sure that **Virginia Tech** is selected as the institution. Click **Login**.
3. The Web-CAT home screen will display useful announcements and assignments that are currently accepting submissions. Find the assignment that you want to submit in the table, and click the "Submit" button next to it.
4. Click the **Browse...** button and select the file you want to upload. The homework assignments and programming projects for this course should be self-contained in a single **.py** file, so you can simply select that one file.
5. Click the **Upload Submission** button. The next page will ask you to review your selection to ensure that you have chosen the right file. If everything looks correct, click **Confirm**.

The next page will show that your assignment is currently queued for grading, with an estimated wait time. This page will refresh itself automatically, and when grading is complete you will be taken to a page with your results.

Pledge

Each of your program submissions must be pledged to conform to the Honor Code requirements for this course. Specifically, you **must** include the following pledge statement in the submitted file:

```
# <include a description of the purpose of this file/project/package>
```

```
#
# @author <name and surname> (your VT PID)
# @date   <the date>
#
# Virginia Tech Honor Code Pledge
# On my honor:
#
# - I have not discussed the Python language code in my program with
#   anyone other than my instructor or the teaching assistants
#   assigned to this course.
# - I have not used Python language code obtained from another student,
#   or any other unauthorized source, either modified or unmodified.
# - If any Python language code or documentation used in my program
#   was obtained from another source, such as a text book or course
#   notes, that has been clearly noted with a proper citation in
#   the comments of my program.
# - I have not designed this program in such a way as to defeat or
#   interfere with the normal operation of the Web-Cat Server.
#
# <your name>
```

Failure to include this pledge in a submission will result in the submission being disallowed during code review.